

Nebraska Department of Education

2008 Statewide Writing Assessment for Grades 4, 8 and 11

Scoring Report

Introduction

The 2008 Nebraska Statewide Writing Assessment was administered in Nebraska public schools to students in grades 4, 8 and 11. A total of 61,243 students were assessed with 20,157 students in grade 4; 20,484 in grade 8; and 20,602 in grade 11. All writing assessments were scored at Educational Service Unit #3, Omaha, NE by scorers from across Nebraska. Grade 4 assessments were scored March 13-15; grade 8, March 27-29; and grade 11, April 3-5.

The purpose of this report is to document the procedures and results of the scoring.

Background

The Statewide Writing Assessment is administered in compliance with State Statute 79-760, which requires that ...“The state board shall prescribe statewide assessments of writing beginning in the spring of 2001 with students in each of three grades selected by the board. For each academic year, one of the three selected grades shall participate in the statewide writing assessment.”

The State Board of Education in 2003 approved the assessment annually of students in grades 4, 8 and 11 beginning in 2004 to meet the federal No Child Left Behind (NCLB) requirements.

The content of the 2008 assessment included a single prompt for each grade level for which students composed a writing sample. Students in grade 4 wrote in the narrative

mode; students in grade 8, the descriptive mode; and students in grade 11, the persuasive mode. The assessment for all grades was administered in a 40-minute session on each of two consecutive days. The writing samples for all grades were scored on a 10-point holistic scale (1, 1+, 2-, 2, 2+, 3-, 3, 3+, 4-, 4). Writing samples for all grades were read and scored by two raters. Raters used a rubric, or scoring guide, that defines performance across six traits (ideas, organization, voice, word choice, sentence fluency, conventions).

Scoring Personnel

Scoring personnel for the 2008 Statewide Writing Assessment included 130 raters for grade 4, 131 raters for grade 8, and 131 raters for grade 11. The raters represented Nebraska geographically as well as from all sizes of school districts. Approximately 54 potential raters were placed on a waiting list. In addition to raters, 5 scoring room leaders and 13 table leaders for each grade level participated in the scoring process. Scoring personnel across all grade levels included a total of 392 raters, 15 scoring room leaders, and 39 table leaders.

Statewide Writing Assessment raters in 2008 were Nebraska educators who were currently teaching at or near the grade levels assessed and/or who were familiar with student writing at grades 4, 8 and 11. Sixty-three percent of raters across all grade levels were currently teaching, while 37% were substitute teachers, retired teachers or educators assigned to other school related positions such as district or building administrator, curriculum specialist, guidance counselor, or media specialist.

Among raters there was a range of years of teaching experience with 16% having 1-5 years of experience, 17% with 6-10 years of experience, 17% with 11-19 years of experience, and 50% with 20 or more years of experience. Raters' years of teaching

experience by grade level are included in Table 1.

Table 1: Raters' Years of Teaching Experience by Grade Level Group

Grade Level	1-5 Years	6-10 Years	11-19 Years	20 Years or More
Grade 4	15%	18%	19%	48%
Grade 8	16%	15%	19%	50%
Grade 11	16%	18%	15%	51%
Average of All Grade Levels	16%	17%	17%	50%

Across all grade levels a majority of raters, 79%, had participated in previous Statewide Writing Assessment scoring sessions. Of the raters who had previously participated in state level scoring, 46% participated in 2006, and 35% in 2007.

A majority of raters across all grade levels, 82%, reported previous experience within their school districts or at regional Educational Service Units in large scale writing assessment scoring. A majority, 94%, reported having participated in some form of training in the six-trait writing assessment model prior to the 2008 state scoring.

Training Procedures for Scoring Personnel

The primary training of raters occurred on the first day of each grade level scoring session. The content of this training included general scoring procedures and the holistic scoring process. The duration of the training was approximately 3.5 hours and included both video taped and direct presentation of information. Topics included in the video taped portion of the training were: responsibilities and expectation for raters; explanation of scoring materials; using the scoring guide to arrive at a holistic score; scoring procedures; criteria for identifying crisis papers and papers that may be unable to be

scored according to the scoring criteria; and sources of rater bias and objectivity issues. Topics included in the direct presentation of information on scoring procedures were: using anchor papers to establish performance levels; using the scoring guide to determine differences in performance levels; allowable and discrepant scores; and understanding inter-rater reliability. During the primary training raters were also required to successfully complete calibration scoring rounds in which they scored papers which had been previously scored but for which the scores had been masked.

In addition to the primary training that raters received at the beginning of the scoring sessions, scoring room leaders conducted re-training activities throughout the remainder of the scoring sessions. The content of these re-training activities included the reviewing the scoring guides, anchor papers, and scoring of additional blind scored papers for purposes of scoring re-calibration. Successful completion of the re-calibration rounds was required in order for raters to proceed with the regular scoring.

Periodically throughout the scoring sessions read-behind scoring checks for scoring consistency were conducted. In this process, each table leader randomly selected a minimum of three scored papers to check for accuracy for each rater at the tables they supervised.

Scoring room leaders and table leaders received training specific to their responsibilities on the day prior to each scoring session. Their responsibilities were to facilitate and supervise the scoring process and to conduct rater training and re-calibration procedures. Training of scoring room leaders and table leaders for each scoring session was conducted by the Nebraska Department of Education Coordinator for Statewide Writing Assessment.

Scoring Process

The scoring process for the Nebraska Statewide Writing Assessment requires that each paper is read and scored by two raters. The two scores are added together to produce a final score ranging from 2.00 to 8.00. The two scores must agree exactly, be no more than one score point apart (adjacent) on the scoring continuum (i.e. 2+, 3-), or be in the same number family (i.e. 2-, 2, 2+). Scores that differ by more than one score point and are not in the same number family (i.e. 2, 3) are considered to be discrepant. Discrepant scores require that the paper be read and scored by a third rater, a process known as adjudication. The adjudication process is intended to resolve the discrepancy according to one of the following scenarios:

Scenario 1

If the 3rd rater's score is **PRECISELY BETWEEN** the 1st and 2nd raters' scores, the 1st and 2nd scores are eliminated, and the 3rd score is doubled to determine the final score. Example: 1st score = 2 (2.00); 2nd score = 3- (2.67); 3rd score = 2+ (2.33). Final score = 4.66.

Scenario 2

If the 3rd rater's score equals either the 1st or 2nd rater's score, the sum of the two equal scores determines the final score. Example: 1st score = 2 (2.00); 2nd score = 3- (2.67); 3rd score = 2 (2.00). Final score = 4.00.

Scenario 3

If the 3rd rater's score is between the 1st and 2nd raters' scores **AND IS ADJACENT** to either the 1st or 2nd score, the 3rd rater's score and the adjacent score are

added together to determine the final score. Example: 1st score = 2 (2.00); 2nd score = 3 (3.00); 3rd score = 2+ (2.33). Final score = 4.33.

Scenario 4

If the 3rd rater's score is between the 1st and 2nd raters' scores BUT IS NOT ADJACENT to either, the 1st and 2nd scores are eliminated, and the 3rd score is doubled to determine the final score. Example: 1st score = 2 (2.00); 2nd score = 4- (3.67); 3rd score = 3 (3.00). Final score = 6.00.

Scenario 5

If the 3rd rater's score is outside the range of the 1st and 2nd raters' scores and is not adjacent to either, the paper is re-scored.

Scenario 6

If the 3rd rater's score is outside the range of the 1st and 2nd raters' scores but is adjacent to either the 1st or 2nd rater score, the non adjacent score is eliminated, and the 3rd rater's score and the adjacent score are added together to determine the final score. Example: 1st score = 2 (2.00); 2nd score = 3 (3.00); 3rd score = 3+ (3.33). Final score = 6.33.

Reliability of the Scoring Process

The reliability of the scoring process is measured by the percentage of agreement among raters in assigning allowable scores. Allowable scores for this scoring process include exact match scores, adjacent scores, and same number family scores. Agreement percentages are based on the scores from first and second ratings only. Percentages of rater agreement for each grade level are included in Table 2.

Table 2: Percentage of Rater Agreement by Grade Level

<u>Grade Level</u>	<u>Exact Match</u>	<u>Adjacent/ Same Family</u>	<u>Overall Agreement</u>	<u>Adjudicated Papers</u>	<u>Papers Scored</u>
Grade 4	45%	51%	97%	3%	20,157
Grade 8	47%	49%	96%	4%	20,484
Grade 11	47%	49%	96%	4%	20,602

Mean Score and Standard Deviation

The range of possible scores on the assessment is from 2.00 to 8.00. The mean or average score for each grade level is included in Table 3. The standard deviation represents the average distance of any given score from the mean or average score. The standard deviation for the scores for each grade level is also included in Table 3.

Table 3: Mean Score and Standard Deviation

<u>Grade Level</u>	<u>Mean Score</u>	<u>Standard Deviation</u>
Grade 4	5.36	1.13
Grade 8	5.82	1.16
Grade 11	5.92	1.10

Process for Appealing Scores

A formal appeals process related to the scoring of the Statewide Writing Assessment was instituted in 2007. Schools wishing to challenge individual student scores as assigned during the state scoring were required to submit a written request for a re-review of a student paper, which included a rationale for the appeal. In all, 43 student writing assessments were submitted for an appeal of the score, including 24 for grade 4, 19 for grade 8, and 0 for grade 11. Papers submitted on appeal were re-reviewed by a

panel of experienced Nebraska educators who had been trained in the Statewide Writing Assessment process and who had participated in the state scoring sessions not less than four years. The results of this appeals process are included in Table 4.

Table 4: Papers Submitted for an Appeal of Scores

<u>Grade Level</u>	<u>Number of Papers Submitted</u>	<u>Papers Receiving Reassigned Scores</u>
Grade 4	24	21
Grade 8	19	11
Grade 11	0	0
<u>Total</u>	<u>43</u>	<u>32</u>

External Scoring of a Sample

A sample of 500 papers from each grade level, or a total of 1,500, was also scored by Measured Progress, Inc., Dover NH, May 13-15, 2008. Papers included in the sample were randomly selected from school districts representing all sizes and geographic regions of the state. The purpose of the external scoring was to validate the in-state scoring processes and procedures.

Scoring Personnel for the Scoring of the Sample

Professional raters at Measured Progress, Inc., who were assigned to score the sample of Nebraska writing assessments, are required as a condition of their employment to hold at least a Bachelor's degree from an accredited college or university and to demonstrate continued proficiency and accuracy in scoring related to the requirements of the various types of assessments they are assigned to score. Raters' previous educational experiences ranged from elementary to post secondary teaching, enrollment in programs of graduate study; and employed in fields related to assessment or education. These

individuals score a variety of assessments for various educational entities and agencies.

Raters were familiar with the six traits of writing as a writing assessment model.

Training Procedures for Scoring Personnel

The Nebraska Department of Education provided the training for the 12 raters at Measured Progress, Inc. following the same procedures and using the same materials as were provided for the in-state scoring training. The same group of raters at Measured Progress, Inc. scored all papers consecutively by grade level beginning with Grade 4.

Reliability of the External Scoring of the Sample

The percentage of agreement among raters in the scoring of the sample was based on the numbers of papers receiving allowable scores assigned on first and second ratings only. Allowable scores for the scoring of the sample included exact match scores, adjacent scores, and same number family scores. Percentages of rater agreement for the scoring of the sample for each grade level are included in Table 5.

Table 5: Percentage of Rater Agreement by Grade Level for Scoring of the Sample

<u>Grade Level</u>	<u>Exact Match</u>	<u>Adjacent/ Same Family</u>	<u>Overall Agreement</u>	<u>Adjudicated Papers</u>	<u>Papers Scored</u>
Grade 4	32%	49%	80%	20%	500
Grade 8	39%	49%	88%	12%	500
Grade 11	41%	47%	88%	12%	500

Mean Score and Standard Deviation for the Sample

The mean or average score for the sample of 500 papers from each grade level scored by Measured Progress, Inc. is included in Table 6. The standard deviation, the average distance of any given score from the mean or average score, is also included in Table 6.

Table 6: Mean Score and Standard Deviation for the Sample

<u>Grade Level</u>	<u>Mean Score</u>	<u>Standard Deviation</u>
Grade 4	5.30	1.38
Grade 8	5.99	1.20
Grade 11	6.13	1.11

English/Spanish Translation for English Language Learners

In 2008 an English/Spanish version of the Nebraska Statewide Writing Assessment for grades 4, 8, and 11 was available as an accommodation or alternate method of assessment for English Language Learner (ELL) students. The translated version is considered an accommodation if the student response is written in English and the response is submitted to the state scoring site. The translated version is considered an alternate method of assessment if the student response is written in a language other than English and the response is scored locally. Districts submitted a request for assessment materials to the NDE by January 14, 2008. Assessment materials were sent to those districts requesting the assessment translation in PDF format on January 18, 2008. Districts printed assessment booklets as needed and administered the assessment within the two-week assessment window.

Translation and Assessment Format

The assessment was translated by Iverson Language Associates, Inc. of Wisconsin. Iverson has over 20 years of experience in the translation industry. Its staff of translators must have industry and translation experience, and must be native speakers of

the target language. The assessment booklet was available in the side-by-side format; English on one page and the Spanish translation on the facing page.

Participation

Twenty-two districts (nine percent) submitted requests for this initial state-issued translation of the writing assessment. Districts indicated the number of students expected to participate in the Statewide Writing Assessment using the English/Spanish version. A total of 771 students participated using the translated version; 341 in grade 4, 223 in grade 8, and 207 in grade 11. This represents ten percent of the state's ELL population in grades 4, 8, and 11. The reason for small numbers might be because participation was an option. The small number might also be the result of communication; notification and information was shared with district contacts via e-mail and Updates.

Survey Results

A brief electronic survey accompanied the assessment materials sent to district assessment contacts. Fifty-nine percent of the districts requesting the English/Spanish version of the assessment responded to the survey.

Districts were asked to identify the language most often used by students to write the response. Twenty-eight percent of the districts identified their student responses were written most often in English, eight percent of the districts identified the number of responses written in English and the number of those written in Spanish was the same, and 23% of the districts identified their student responses were written most often in Spanish.

Seventy-seven percent perceived the translated version of the writing assessment as helpful to their ELL students, 84% responded that the state-issued translation should

be available in the future, and 84% would recommend their district use the translated version. One district responded that it did not find the translated version helpful, would not recommend the district avail itself of the translated version, but did not have an opinion about whether the state-issued translation should be available in the future. One district did not respond to these questions.

Summary and Conclusion

Scoring personnel for the 2008 Nebraska Statewide Writing Assessment for grades 4, 8 and 11 included a total of 392 raters. Across all grade levels a majority of raters were currently teaching in Nebraska classrooms and were experienced educators with 11 or more years of teaching experience. A majority of raters also reported previous experience in large scale writing assessment scoring at the state, regional and local levels as well as participation in six trait writing assessment training prior to the 2008 state scoring.

In addition to the in-state scoring of the assessment, a sample of 500 papers for each grade level, or a total of 1,500 papers, was scored externally. The purpose of the external scoring of the sample was to validate the processes and procedures used for the in-state scoring.

An analysis of the mean scores for the papers scored in-state and the externally scored sample shows that the in-state mean scores for grades 8 and 11 are lower than the corresponding mean scores for the sample for grades 8 and 11. The in-state mean for grade 4 was higher than the corresponding mean score for the sample for grade 4. The in-state mean score for grade 4 is .06 higher than the corresponding mean score for the sample for grade 4. The standard deviations for papers scored in-state across all grade

levels are lower than the corresponding standard deviations for the externally scored sample across all grade levels.

The percentages of overall rater agreement, however, were significantly higher, across all grade levels for the papers scored in-state than for papers scored in the externally scored sample. This result may indicate that raters of the in-state scoring were significantly more consistent in their understanding and application of the criteria of the scoring guides to the scoring process than raters who scored the sample. This result may also be indicative of the fact that the majority of the raters for the in-state scoring were currently teaching, and were experienced in number of years of classroom teaching and in their understanding of the six trait writing process. None of the raters for the externally scored sample were currently teaching and may or may not have had previous teaching experience or experience in the six trait writing model beyond their work as professional assessment raters.

An English/Spanish version of the Nebraska Statewide Writing Assessment was available as an accommodation or alternate method of assessment for Spanish-speaking English Language Learners. Districts requested and were provided the translated assessment materials electronically. The translated version was administered during the required assessment window, and student responses written in English were submitted for scoring to the state scoring site. Responses written in a language other than English were scored locally. Nine percent of the districts availed themselves of the translated materials. The majority felt the English/Spanish version was helpful to ELL students and would recommend continued use.